Nucleic Acids Research, 1993, Vol. 21, No. 10 **2383–2388**

Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*

E.Jean Finnegan and Elizabeth S.Dennis
CSIRO, Division of Plant Industry, Canberra, ACT 2601, Australia

Received February 17, 1993; Revised and Accepted April 15, 1993 GenBank accession no. L10692

ABSTRACT

A plant cytosine methyltransferase cDNA was isolated using degenerate oligonucleotides, based on homology between prokaryote and mouse methyltransferases, and PCR to amplify a short fragment of a methyltransferase gene. A fragment of the predicted size was amplified from genomic DNA from *Arabidopsis thaliana*. Overlapping cDNA clones, some with homology to the PCR amplified fragment, were identified and sequenced. The assembled nucleic acid sequence is 4720 bp and encodes a protein of 1534 amino acids which has significant homology to prokaryote and mammalian cytosine methyltransferases. Like mammalian methylases, this enzyme has a C terminal methyltransferase domain linked to a second larger domain. The *Arabidopsis* methylase has eight of the ten conserved sequence motifs found in prokaryote cytosine-5 methyltransferases and shows 50% homology to the murine enzyme in the methyltransferase domain. The amino terminal domain is only 24% homologous to the murine enzyme and lacks the zinc binding region that has been found in methyltransferases from both mouse and man. In contrast to mouse where a single methyltransferase gene has been identified, a small multigene family with homology to the region amplified in PCR has been identified in *Arabidopsis thaliana*.

INTRODUCTION

The most common modification of DNA in higher eukaryotes is methylation of cytosine residues at carbon 5. In vertebrates, 3–8% of cytosines are methylated (1) while in plants up to 30% of cytosines are modified (2). The difference in extent of cytosine methylation between vertebrates and plants can be attributed to two factors. DNA methylation in animals is generally confined to cytosines in CG dinucleotides while in plants methylation occurs at cytosines located in both CG dinucleotides and CNG triplets, where N is any base (3). In addition the CG dinucleotide is more common in DNA of plants than of animals. DNA methylation has been implicated in regulating gene expression during development, in determining chromatin structure and in compartmentalization of DNA (reviewed in 4, 5, 6).

Methyl groups are transferred to cytosine residues from S-adenosyl methionine in a reaction catalysed by a DNA methyltransferase or methylase (7). Prokaryote cytosine methyltransferases generally methylate cytosines within a longer target sequence, while mammalian methyltransferases methylate cytosine residues in any CG dinucleotide. Prokaryote cytosine methyltransferases are structurally similar with highly conserved motifs alternating with less well conserved sequences (8, 9, 10). The presence of these conserved motifs, designated I to X, which occur in the same order in all these enzymes differentiates cytosine-5 methyltransferases from cytosine-N4 and adenine-N6 methylases (9). The cysteine residue of a highly conserved proline-cysteine doublet (motif IV, PCXXXS) forms the active-site (11, 12, 13, 14), while motif I contains a sequence that is thought to bind S-adenosyl methionine (F/GXGXG, 15). The target recognition domain which specifies both the target sequence and the base to be methylated lies in the variable region between motifs VIII and IX (12, 16, 17, 18).

Mammalian cytosine methyltransferases are comprised of two protein domains which fold independently, a C terminal methyltransferase domain which is structurally similar to that of prokaryote methylases, fused to a second large domain (19, 20). A single DNA methyltransferase gene has been detected in the mouse which is consistent with the finding that there is a single species of methylase in different cell types which differ in the pattern of DNA methylation (19). Partial purification of a methyltransferase enzyme from pea (21, 22), wheat (23) and rice (24) has been reported but no plant methyltransferase genes have been cloned. Plants differ from animals in both the extent and sequence specificity of methylation. It is not known whether plant methyltransferases resemble the family of prokaryote enzymes each of which recognizes several target sites, or whether plants have multiple methyltransferases which catalyse methylation at CG and CNG respectively.

We have cloned a cytosine methyltransferase gene from *Arabidopsis thaliana*. We used homology between the mouse methyltransferase and prokaryote methylases to design primers for PCR amplification of a fragment spanning conserved motifs IX and X. Taking this approach we have isolated overlapping

clones which make up a full length cDNA for a methyltransferase gene from *Arabidopsis*. The inferred amino acid sequence of this composite cDNA shows homology to cytosine methyltransferases in the C terminal domain. A family of genes, with homology to the amplified region, has been identified in a Southern analysis.

MATERIALS AND METHODS

PCR amplification of fragment of a DNA methylase

Primers for PCR were MMet1 (CCGAATTCCAG/AGGNTTT/CCCNGAC/T) for region IX and MMet2 (CGGGATCCACNGCA/GTTNCCNACC/TTG) for region X (Figure 1a) where N represents a mix of all four bases. Restriction endonucleases recognition sites are indicated in bold.

The final reaction conditions for PCR were 250ng genomic DNA or 1ng linearized plasmid DNA, 1μM each primer, 200μM each dNTP, 10mM Tris pH8.8 at 25°C, 1.5mM $MgCl_2$, 50mM KCl, 0.1% Triton X-100 and 1 unit *Taq* DNA polymerase in a reaction volume of 25μl. Cycling conditions included an initial denaturation step at 95°C for 5 minutes followed by 30 cycles of 46°C for 30 seconds, 65°C for 30 seconds and 95°C for 30 seconds. The final cycle was 46°C, 30 seconds and 65°C for 2 minutes. Reaction products were separated on an 8% polyacrylamide gel and amplified fragments were isolated from the gel, digested with *Bam*HI and *Eco*RI, then cloned into pUC19. Six independent clones were sequenced to eliminate any errors introduced during PCR amplification.

Primers specific for sequences towards the 5'end of clone Yc8 were used to amplify the region spanning the overlap between cDNAs Yc8 and Yc2 (Figure 2). The primer AMet1 (CCTAGACTCTCACCATCCC) was used to prime the reverse transcription of total RNA, the products were tailed in the presence of dATP and then amplified by PCR using AMet1 and an oligodT primer containing recognition sites for the enzymes *Eco*RI, *Sma*I and *Bam*HI at the 5' end. The nested primer AMet2 (GCGGATCCTTCCAGAACTGCCTCGG) was used in a subsequent PCR amplification of 1μl of the initial PCR mix with the same oligodT primer (25). Products of this amplification were gel purified, cleaved with *Bam*HI and cloned into *Bam*HI cleaved pUC19 for sequencing.

Screening of λ genomic and cDNA libraries

The cloned, PCR amplified fragment was gel purified then random primed in the presence of both $dATP^{32}$ and $dCTP^{32}$ (Oliglabelling Kit, Bresatech). A genomic library (Promega) containing *A.thaliana* Landsberg DNA, partially digested with *Mbo*I, was screened with this probe and positive plaques identified. An *Eco*RI fragment, that hybridized to the probe, was subcloned from a positive plaque for further mapping and sequence analysis. A 400bp *Hinc*II/*Eco*RI fragment which encompassed the amplified region was identified and used for screening a cDNA library (Landsberg, Promega). Three overlapping cDNA clones were identified in this way. Screening of a second cDNA library (Columbia with low percent Landsberg, J. Mulligan, pers comm., 26) with the cDNA probe Pc2 (Figure 2), identified more overlapping cDNA clones.

Sequencing and sequence analysis

Templates for sequencing were derived by subcloning and by generating several series of nested deletions by *Exo*III deletion (*Exo*III nested deletion kit, Pharmacia). Sequencing was done in the presence of radiolabelled dATP using T_7 polymerase (T_7 sequencing kit, Pharmacia), or with fluorescent dye labelled primers and *Taq* DNA polymerase (*Taq* dye primer cycle sequencing kit, Applied Biosystems). Sequences were obtained for both strands of at least one cDNA clone, and in some instances were confirmed by sequencing one strand of an overlapping clone. Sequencing reactions using fluorescent dye primers were resolved on an Applied Biosystems 370A DNA Sequencer. Nucleotide and amino acid comparisons were done using GCG7.1 sequence analysis package (27).

Southern hybridization

The probes used for Southern analysis were the PCR amplified fragment, spanning conserved regions IX and X (see Results) and a 398bp fragment from cDNA clone Pc2 that encompasses the region amplified in PCR (71bp), flanked by coding DNA of 84 bp 5' and 78bp 3' and 165bp 3' untranslated (probe 1, Figure 2). DNA used in Southern analyses was isolated from ecotype Landsberg and the hybridization procedure has been described previously (28).

RESULTS

PCR amplification of part of a methyltransferase gene from *Arabidopsis*

The methyltransferase domain of mammalian methylases retains eight of the ten conserved sequence motifs that are characteristic of prokaryote cytosine methyltransferases (29, 9). A comparison of the amino acid sequence for the mouse methylase (19) with methylases M.*Hha*I (30), M.*Dde*I (31) and M.*Eco*RII (8) for motifs IX and X revealed regions of homology with low codon degeneracy (Figure 1a). Degenerate oligonucleotides, corresponding to the mouse amino acid sequence in these regions, were designed to prime amplification of the short variable region between motifs IX and X, that is 71 nucleotides in the mouse gene. Restriction endonuclease recognition sites were included at the 5' end of each primer to facilitate cloning which increased the expected length of a fragment amplified from a methyltransferase gene to 87 bp.

These primers were used in PCR with genomic DNA template from a number of plant species but a band of the predicted size (87 bp) was amplified only from *Arabidopsis* DNA (Figure 1b), possibly because of its small genome size. This fragment was cloned and six independent isolates sequenced. The sequence of five clones was identical while the remaining clone had eight bases inserted between one primer and the rest of the sequence, which was identical to that in the other clones. This was the only clone out of 24 examined that contained a larger insert and may be an artefact of PCR amplification.

The deduced amino acid sequence, excluding residues encoded by the primers, was 54% identical (7/13) to the mouse methyltransferase in this region (Figure 1c) suggesting that the fragment amplified represents a fragment of a plant methyltransferase gene. A number of other bands were also amplified from the *Arabidopsis* DNA template (Figure 1b). Amplification of only two of these was dependent upon the addition of both primers to the reaction mix and sequence analysis showed that these fragments were not homologous to the mouse methyltransferase. The remaining bands were amplified when only one primer, either MMet1 or MMet2, was included in the reaction mix; these bands were not characterized.

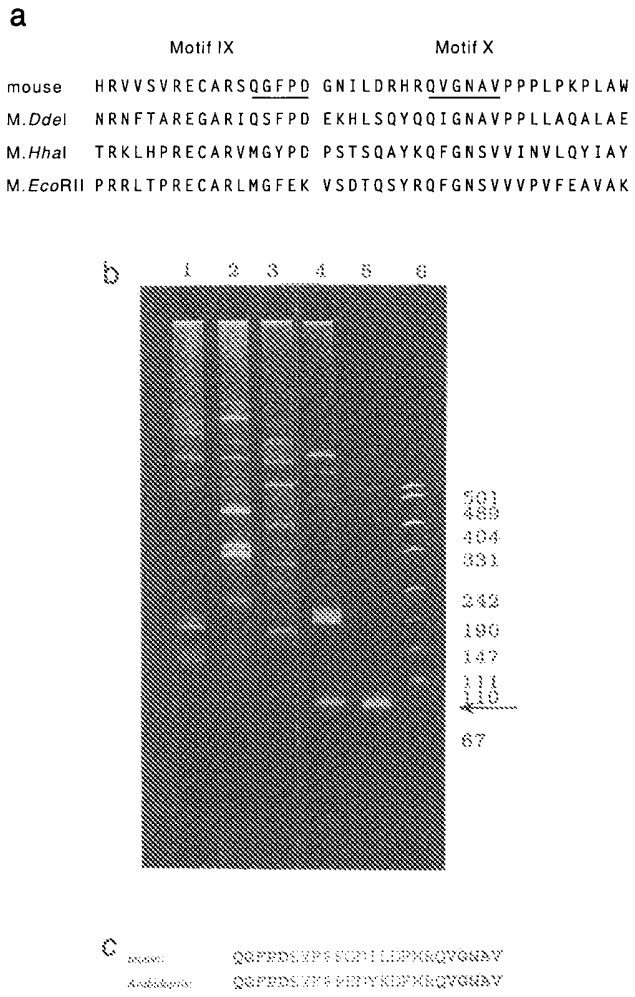

Figure 1a. Motifs IX and X from the mouse methyltransferase (19) and prokaryote methylases M.*Hha*I (30), M.*Dde*I (31) and M.*Eco*RII (8) were compared. Strongly conserved regions with low codon degeneracy were selected for synthesis of oligonucleotide primers for PCR. The sequences represented in the primers MMet1 and MMet2 are underlined. **b.** Acrylamide gel showing the products of PCR amplification using primers MMet1 and MMet2 with genomic DNA template from pea (lane 1), flax (lane 2), cotton (lane 3), *Arabidopsis* (lane 4) and a plasmid containing the mouse methyltransferase cDNA (lane 5). The marker (lane 6) is pUC19 cut with *Hpa*II. The band indicated with an arrow was cloned and sequenced. **c.** Comparison of the deduced amino acid sequence of the fragment amplified from *Arabidopsis* DNA in PCR and the corresponding region of the mouse methylase. The sequences represented by the primers are in bold.

Cloning of an *Arabidopsis* methyltransferase cDNA

The 87 bp PCR amplified product was used to screen a genomic library and positive plaques identified. Sequence analysis of one clone, using the primers MMet1 and MMet2, showed that the nucleotide sequence was identical to that amplified in PCR. A *Hin*cII-*Eco*RI fragment (approximately 400 bp), encompassing the amplified product, was isolated from this clone and used to screen a cDNA library. Three cDNA clones ranging in size from about 300bp to 700bp were identified among approximately 160,000 screened. The largest of these clones, Pc2 (Figure 2), was used to screen a second cDNA library (26). Four positive clones were isolated and the 5′ end of the longest clone, Yc8 (probe 2, Figure 2), was used to rescreen this library to isolate clone Yc2, which extends beyond the end of the coding region

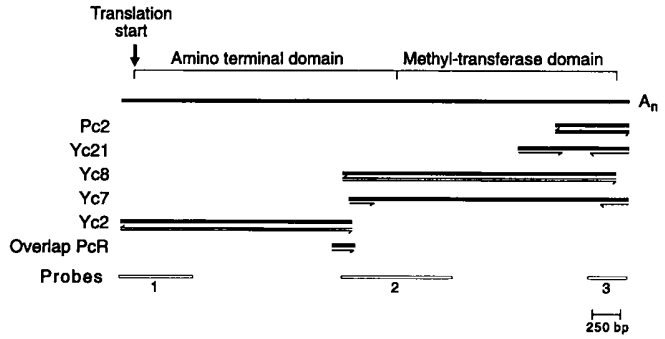

Figure 2. Location of the overlapping cDNA clones and the PCR fragment, used to verify the overlap between clones Yc8 and Yc2, with respect to the methyltransferase and amino terminal domains. The thin lines indicate the regions that were sequenced, the arrowheads indicate the direction of sequencing. The location of probes is also indicated; probe 1 is referred to as the 398 bp probe from Pc2, probe 2 as the 5′ end of clone Yc8 and probe 3 as the Yc2 probe.

(Figure 2). Clones Pc2, Yc8 and Yc2 were completely sequenced on both strands, while only 300–400 bases at each end of clones Yc7 and Yc21 were sequenced (Figure 2).

Clones Yc7 and Yc8 share a common restriction map and the sequence of the coding regions of clones Pc2, Yc8, Yc7 and Yc21, where sequenced (Figure 2), are identical indicating that these clones are derived from the same gene. The 3′ untranslated regions of clones Pc2, Yc7 and Yc21 are also identical in sequence but differ in length preceding the polyA tail. The sequence of clone Yc8 diverges from that of the other clones 49 bp beyond the stop codon. This sequence difference may arise because these clones are from different alleles or may reflect a difference between ecotypes Columbia and Landsberg, both represented in the λYES library (J. Mulligan, pers comm.).

The overlap between Yc8 and Yc2 was confirmed by sequencing 4 independent isolates of a fragment amplified from the products of a first strand cDNA synthesis using nested methylase specific primers (Figure 2). Additional confirmation that these clones are derived from the same gene comes from the isolation of a genomic clone that hybridizes to both Pc2 and Yc2 (probes 1 and 3, Figure 2). The length of the methyltransferase cDNA assembled from the overlapping cDNA clones Yc8 and Yc2 is 4720bp not including a poly A tail (Accession No. L10692), which agrees with the estimate based on Northern analysis of 4.7kb (data not shown).

The assembled nucleotide sequence encodes an open reading frame of 1534 amino acids, comparable in length to the murine enzyme (1587 aa). There is an inframe stop codon 66 bases upstream of the first methionine. The cDNA differs from the PCR amplified product (70% homology at the amino acid level) in both the region between the primers and in one amino acid of the priming site in region X, resulting in a mismatch at the fourth base from the 3′ end of primer MMet2. This mismatch could account for the failure to detect this sequence amongst the six PCR clones that were originally characterized. The difference in sequence between the PCR amplified product and the cDNA clones suggests that they may represent different genes.

Sequence comparison with the mouse methyltransferase

The inferred amino acid sequences of the *Arabidopsis* and mouse enzymes are 50% homologous in the C terminal methyltransferase domain (Figure 3). The eight motifs conserved

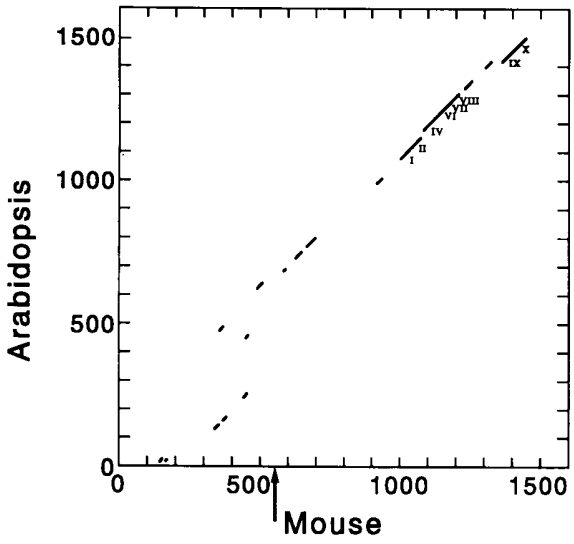

Figure 3. A dotplot comparison between the mouse (horizontal axis) and *Arabidopsis* (vertical axis) methyltransferase proteins. The window size is 30 and stringency of the match is 18. The conserved sequence motifs in the methyltransferase domain are indicated by numbers I to X beside the diagonal. The arrow head indicates the location of the zinc binding region in the mouse enzyme.

MOUSE:	EMLCGGPPCQGFSGMN
HUMAN:	EMLCGGPPCQGFSGMN
ARABIDOPSIS:	DFINGGPPCQGFSGMN
PROKARYOTE CONSENSUS:	D----G-PCP-FS--G N---------Q-W----

Figure 4. Comparison of conserved motif IV which contains the active cysteine residue in cytosine-5 prokaryote methyltransferases. In the prokaryote consensus sequence alternate amino acids are listed one above the other and variable residues are indicated by a dash.

in all eukaryote and prokaryote cytosine methyltransferases are present in the same order in the plant methylase as in both prokaryote and the other eukaryote enzymes (Figures 3). A proline-cysteine doublet present in conserved motif IV has been identified as the functional catalytic domain in prokaryotic cytosine-5 methyltransferases (11, 12, 13, 14). This motif is highly conserved between the prokaryote, mammalian and plant enzymes (Figure 4), suggesting that the prolyl-cysteinyl doublet may also be the catalytic site in eukaryote enzymes. The S-adenosyl methionine binding domain is also conserved in the plant enzyme (motif I, Figure 3).

The variable region between conserved motifs VIII and IX determines the sequence specificity of methylation in the prokaryote methylases (12, 16, 17, 18, 32). Prokaryote methyltransferases that recognize identical or similar target sequences have homology in this region while enzymes recognizing different targets show little or no homology (9, 33, 34). The mouse and *Arabidopsis* proteins also have homology in this region, but it does not extend the full length of the target recognition domain (Figure 3). In addition there is a deletion of

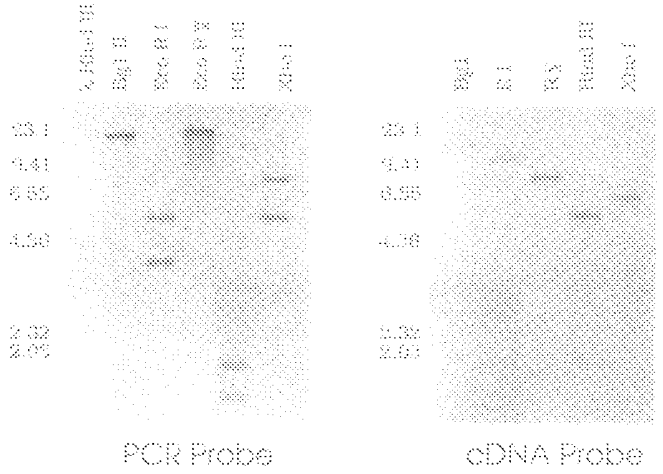

Figure 5a. Southern hybridization of *Arabidopsis* DNA cleaved with enzymes as indicated and probed with the PCR amplified fragment that encodes 71 bases of coding sequence for part of conserved motifs IX and X and the variable region between these motifs. **b.** Southern hybridization of the same filter as shown in Figure 5a, probed with a 398 bp fragment of cDNA Pc2 that has homology to the PCR amplified fragment, plus 84 bp (5′) and 78 bp (3′) coding sequence and 165 bp of 3′ untranslated sequence (probe 1, Figure 2).

44aa from the *Arabidopsis* protein towards the C terminal end of this region.

The amino terminal domain of the mouse and human enzymes is separated from the methyltransferase domain by 13 alternating lysine and glycine residues. These domains can be separated by proteolytic cleavage suggesting that they fold independently (20, 35). The methyltransferase domain of the *Arabidopsis* protein is separated from the amino terminal domain by the sequence KKKGKG. While this differs from the corresponding sequence in the mammalian enzymes it is also lysine rich, suggesting that it may have some functional significance.

The most striking feature of the amino terminal domain is the relative lack of homology (24%) between the mouse and *Arabidopsis* proteins (Figure 3). This contrasts to the mouse and human methyltransferases which are 70% identical in this domain compared to 83% in the methylase domain (29). Homology between the mouse and *Arabidopsis* proteins is limited to short stretches throughout this domain (Figure 3). One of these regions of homology lies between residues 300−450 in the mouse enzyme; residues 207−455 target the mouse methyltransferase to replication forks in S phase nuclei (36). A zinc binding domain, $CX_2CX_2CX_4CX_2CX_2CX_{15}CX_4C$, has been identified within the N terminal domain of the mouse protein (35). This domain is conserved in the human enzyme (29), but does not occur in the *Arabidopsis* protein. An acidic region, which contains 16 glutamic acid and two aspartic acid residues in 36 aa (residues 656 to 692, Figure 3) found in the plant enzyme is not present in the mammalian enzymes. A sequence which resembles recognized nuclear localization signals (reviewed in 37) occurs near the amino terminus of the plant enzyme.

Identification of a methyltransferase gene family in *Arabidopsis*

Partial purification of the methyltransferase enzyme from pea has failed to identify two distinct methyltransferase functions, one specific for CG and a second for CNG (21, 22). However, in

tobacco, methylation of CG and CNG motifs occurring in repeated sequences showed different sensitivity to the inhibitor ethionine (38). While 5-azacytidine treatment resulted in demethylation of cytosines in both CG and CNG motifs, ethionine treatment caused marked demethylation of CNG triplets with little effect on CG methylation (38). Ethionine may alter the specificity of a single methyltransferase (38), or CG and CNG motifs may be methylated by separate enzymes which differ in their sensitivity to ethionine.

Southern analyses of *Arabidopsis* DNA using the PCR amplified fragment (71bp coding sequence) showed a single band when DNA was cut with some enzymes (*Bgl*II and *Eco*RV) and two bands when DNA was cleaved with *Eco*RI, *Hin*dIII or *Xho*I (Figure 5a). This indicates that there are two copies of this region in the genome because the fragment used as a probe was amplified directly from genomic DNA and does not encode sites for the enzymes *Eco*RI, *Hin*dIII or *Xho*I. The presence of a single large band when DNA was cut with *Bgl*II and *Eco*RV suggests that these copies may be linked. We have identified a single Yac (39) that encodes both copies of this fragment on an 80kb fragment of *Arabidopsis* DNA which supports the idea of linkage (data not shown).

The same blot was reprobed with a 398 bp fragment from cDNA clone Pc2 (Figure 5b). This probe encodes a region 86% homologous (over 71bp) to the PCR amplified region flanked by 84 bp (5′) and 78 bp (3′) of coding sequence and 165 bp of 3′ untranslated sequence including a polyA tail (18 b) (probe 1, Figure 2). At high stringency one strongly hybridizing band is seen in each lane plus a second band of much lower intensity; with the exception of DNA cut with *Bgl*II where there is one band in common, neither of these bands comigrates with the bands identified by the PCR probe. After extended exposure of the autoradiogram, or when the hybridization stringency was reduced, the bands identified by the PCR probe can be detected by hybridization to the cDNA probe (data not shown). This suggests that there is a small family of genes with homology to a DNA methyltransferase.

DISCUSSION

We suggest that the inferred amino acid sequence of the *Arabidopsis* protein described in this report is a cytosine methyltransferase based on its homology to both mammalian and prokaryotic cytosine-5 methyltransferases in the C terminal or methyltransferase domain (9, 19, 29). Like the mammalian enzymes, the *Arabidopsis* enzyme has eight of the ten regions characteristic of the prokaryote cytosine methyltransferases (9). Motifs I and IV, which have been identified as the S-adenosyl methionine binding domain and the active site respectively, are highly conserved between this *Arabidopsis* enzyme, both mammalian methyltransferases and all prokaryote cytosine methyltransferases. The presence of the 8 motifs, found only in cytosine-5 methyltransferases, is strong evidence that the *Arabidopsis* protein described here also functions as a DNA methyltransferase.

The variable target recognition domain between motifs VIII and IX is less well conserved, between the plant and mammalian enzymes, than the remainder of the methylase domain. Homology between the mouse and *Arabidopsis* proteins in this region suggests that they may share a common target sequence, that is CG dinucleotides, but the observed differences makes this less than certain.

In both mammalian enzymes, the N terminal domain is separated from the methyltransferase domain by a run of alternating lysine-glycine residues. There is a shorter glycine-lysine rich sequence separating the two domains in the *Arabidopsis* protein. Although the homology between the mouse and human proteins is somewhat lower in the N-terminal domain than in the methyltransferase domain (70% compared to 83%), the two proteins are still highly conserved. In contrast the mammalian and plant enzymes show only 24% homology in the N-terminal domain; homologous regions are short and scattered throughout this domain. Perhaps significantly, the region (residues 207−455) that targets the mouse enzyme to the replication fork in S phase nuclei (36) shows homology to residues 120−280 in the *Arabidopsis* protein, suggesting that the latter may also be located at the replication fork. While no function has been assigned to other regions of homology, conservation of these sequences between plants and mammals suggests that they may be essential for enzyme function. The most significant feature of the mammalian enzymes in this domain, a zinc binding region, is absent from the *Arabidopsis* protein. The motif S/TPXX, where X tends to be a basic amino acid, occurs frequently in regulatory proteins; these motifs bind in the minor groove of DNA with the narrower minor groove of AT rich DNA being the preferred binding site (40). The presence of this motif, which occurs ten times in the N terminal domain of the mouse methyltransferase (41) and five times in the plant enzyme, may indicate this domain is involved in binding DNA.

The two protein domains of the mouse enzyme fold independently and can be separated by proteolytic cleavage. When the N terminal domain was cleaved from the methyltransferase domain the latter retained activity; separation of the two domains caused a large stimulation in the rate of *de novo* methylation, that is methylation of unmethylated DNA (35). The rate of methylation of a hemimethylated substrate was not significantly changed by separation of the two domains, suggesting that the amino terminal domain down regulates *de novo* methylation by the intact enzyme (35). It should now be possible to determine the function of the corresponding domain in the plant enzyme. Bacterial methylases, which have no counterpart to the amino terminal domain, show no discrimination between unmethylated and hemimethylated DNA. The mammalian methyltransferases may have arisen by fusion of two ancestral genes, one with methyltransferase activity and the other a sequence specific DNA binding protein (20, 35). The finding that the *Arabidopsis* methyltransferase lacks the zinc binding domain and shows only limited homology to the mouse protein in the amino terminal domain suggests that this domain has evolved more rapidly than the methyltransferase domain. Alternatively, gene fusion giving rise to the complex methyltransferase in eukaryotes may have occurred independently in the plant and animal kingdoms.

In contrast to the mouse where a single methyltransferase gene has been detected (19), a small multigene family with homology to the region amplified in PCR (regions IX−X) has been identified in *Arabidopsis* DNA. At least two members of this family (described above), genes represented either by cDNA clones or by the PCR amplified product and its corresponding genomic clone, have greatest homology in a data base search to other cytosine-5 methyltransferases. This gene family may encode enzymes that differ in specificity of methylation, for example methylating cytosines in CG or CNG motifs, in the time during development at which they are expressed, or which are targeted to the chloroplast rather than the nucleus. Studies with

transgenic plants will clarify the function and regulation of these genes.

ACKNOWLEDGEMENTS

The authors thank Tim Bestor for his gift of the mouse methyltransferase cDNA clone and for helpful discussions, Stephen Elledge for his gift of an *Arabidopsis* λ cDNA library, Erwin Grill for his gift of an *Arabidopsis* YAC library and Jim Peacock for critical discussions. Lyndall Thorpe provided excellent technical assistance. This work was supported, in part, by the 'May Statement Initiative' funding for 'Geneshears'.

REFERENCES

1. Shapiro, H.S. (1975) Content of 6-methylaminopurine and 5-methylcytosine in DNA. CRC Press. Boca Raton, FL.
2. Adams, R.L.P. and Burdon, R.H. (1985) Molecular biology of DNA methylation. Springer Verlag. New York.
3. Gruenbaum, Y., Naveh-Many, T., Cedar, H. and Razin, A. (1981) *Nature*, **292**, 860–862.
4. Razin, A. and Cedar, H. (1991) *Microbiolog. Rev.*, **55**, 451–458.
5. Lewis, J. and Bird, A. (1991) *FEBS Lett*, **285**, 155–159.
6. Finnegan, E.J., Brettell, R.I.S. and Dennis, E.S. (1993) In Jost, J.P. and Saluz, H. P. (eds), DNA methylation: Molecular biology and biological significance. Birkhäuser Verlag, Basel, pp 218–261.
7. Adams, R.L.P. and Burdon, R.H. (1983) In Jacob, S.T. (ed.), Enzymes of nucleic acid synthesis and processing. CRC Press, Boca Raton, FL, 119–144.
8. Som, S., Bhagwat, A.S. and Friedman, S. (1987) *Nucl. Acids Res.*, **15**, 313–331.
9. Pósfai, J., Bhagwat, A.S., Pósfai, G. and Roberts, R.J. (1989) *Nucl. Acids Res.*, **17**, 2421–2435.
10. Lauster, R., Trautner, T.A. and Noyer-Weidner, M. (1989) *J. Mol. Biol.*, **206**, 305–312
11. Wu, J.C. and Santi, D.V. (1987) *J. Biol. Chem.*, **262**, 4778–4786.
12. Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B. and Trautner, T.A. (1988) *EMBO J.*, **7**, 2601–2609.
13. Chen, L., MacMillan, A.M., Chang, W., Ezaz-Nikpay, K., Lane, W.S. and Verdine, G.L. (1991) *Biochemistry*, **30**, 11018–11025.
14. Wyszynski, M.W., Gabbara, S. and Bhagwat, A.S. (1992) *Nucl. Acids Res.*, **20**, 319–326.
15. Ingrosso, D., Fowler, A.V., Bleibaum, J. and Clarke, S. (1989) *J. Biol. Chem.*, **264**, 20131–20139.
16. Balganesh, T.S., Reiners, L., Lauster, R., Noyer-Weidner, M., Wilke, K. and Trautner, T.A. (1987) *EMBO J.*, **6**, 3543–3549.
17. Trautner, T.A., Balganesh, T.S. and Pawlek, B. (1988) *Nucl. Acids Res.*, **16**, 6649–6658.
18. Klimasauskas, S., Nelson, J.L. and Roberts, R.J. (1991) *Nucl. Acids Res.*, **19**, 6183–6190.
19. Bestor, T., Laudano, A., Mattaliano, R. and Ingram, V. (1988) *J. Mol. Biol.*, **203**, 971–983.
20. Bestor, T.H. (1990) *Phil. Trans. R. Soc. Lond. B.*, **326**, 179–187.
21. Yesufu, H.M.I., Hanley, A., Rinaldi, A. and Adams, R.L.P. (1991) *Biochem. J.*, **273**, 469–475.
22. Adams, R.L.P., Bryans, M., Rinaldi, A., Smart, A. and Yesufu, H.M.I. (1990) *Phil Trans. R. Soc. Lond.B.*, **326**, 189–198.
23. Theiss, G., Schleicher, R., Schimpff-Weiland, R. and Follman, H. (1987) *Eur. J. Biochem.*, **167**, 89–96.
24. Giordano, M., Mattachini, M.E., Cella, R. and Pedrali-Noy, G. (1991) *Biochem. Biophys. Res. Comm.*, **177**, 711–719.
25. Frohman, M.A. (1990) In Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J.(eds.), PCR protocols. Academic Press, 28–38.
26. Elledge, S.J., Mulligan, J.T., Ramer, S.W., Spottswood, M. and Davis, R.W. (1991) *Proc. Natl. Acad. Sci. USA.*, **88**, 1731–1735.
27. Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucl. Acids Res.*, **12**, 387–395.
28. Taylor, B.H., Finnegan, E.J., Dennis, E.S. and Peacock, W.J. (1989) *Plant Mol. Biol.*, **13**, 109–118.
29. Yen, R.-W.C., Vertino, P.M., Nelkin, B.D., Yu, J.J., El-Deiry, W., Cumaraswamy, A., Lennon, G.G., Trask, B.J., Celano, P. and Baylin, S.B. (1992) *Nucl. Acids Res.*, **20**, 2287–2291.
30. Caserta, M., Zacharias, W., Nwanko, D., Wilson, G. and Wells, R. (1987) *J. Biol. Chem.*, **262**, 4770–4777.
31. Sznyter, L., Slatko, B., Moran, L., O'Donnell, K. and Brooks, J. (1987) *Nucl. Acids Res.*, **15**, 8249–8266.
32. Mi, S. and Roberts, R.J. (1992) *Nucl. Acids Res.*, **20**, 4811–4816.
33. Walter, J., Noyer-Weidner, M. and Trautner, T.A. (1990) *EMBO J.*, **9**, 1007–1013.
34. Szilak, L., Venetianer, P. and Kiss, A. (1990) *Nucl. Acids Res.*, **18**, 4659–4664.
35. Bestor, T.H. (1992) *EMBO J.*, **11**, 2611–2617.
36. Leonhardt, H., Page, A. W., Weier, H.-U., Bestor, T. H. (1992). *Cell* **71**, 865–873.
37. Garcia-Bustos, J., Heitman, J. and Hall, M.N. (1991) *Biochim. Biophys. Acta.*, **1071**, 83–101.
38. Bezdek, M., Koukalova, B., Kuhrova, V. and Vyskot, B. (1992) *FEBS Letters.*, **300**, 268–270.
39. Grill, E. and Sommerville, C.(1991) *Mol. Gen. Genet.*, **226**, 484–490.
40. Churchill, M.E.A. and Suzuki, M. (1989) *EMBO J.*, **8**, 4189–4195
41. Bestor, T.H., Gundersen, G., Kolstø, A.-B. and Prydz, H.(1992) *Genetic Analysis—Techniques and Applications.*, **9**, 48–53.